THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 2002 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 277, No. 47, Issue of November 22, pp. 45108–45114, 2002
Printed in U.S.A.

Functional Consequences of Insertions and Deletions in the Complementarity-determining Regions of Human Antibodies*

Received for publication, August 16, 2002, and in revised form, September 16, 2002
Published, JBC Papers in Press, September 16, 2002, DOI 10.1074/jbc.M208401200

Johan Lantto and Mats Ohlin‡

From the Department of Immunotechnology, Lund University, P.O. Box 7031, S-220 07 Lund, Sweden

Insertions and deletions of nucleotides in the genes encoding the variable domains of antibodies are natural components of the hypermutation process, which may expand the available repertoire of hypervariable loop lengths and conformations. Although insertion of amino acids has also been utilized in antibody engineering, little is known about the functional consequences of such modifications. To investigate this further, we have introduced single-codon insertions and deletions as well as more complex modifications in the complementarity-determining regions of human antibody fragments with different specificities. Our results demonstrate that single amino acid insertions and deletions are generally well tolerated and permit production of stably folded proteins, often with retained antigen recognition, despite the fact that the thus modified loops carry amino acids that are disallowed at key residue positions in canonical loops of the corresponding length or are of a length not associated with a known canonical structure. We have thus shown that single-codon insertions and deletions can efficiently be utilized to expand structure and sequence space of the antigen-binding site beyond what is encoded by the germline gene repertoire.

Antibodies are highly specific receptors of the immune system that also have a great potential as reagents in biological chemistry and as therapeutic agents. The part of the antibody that makes contact with the antigen is comprised of two variable (V^1 domains, the heavy (H) and the light (L), which both are made up of a two-β-sheet framework. From this framework, six complementarity-determining region (CDR) loops, three from the light domain and three from the heavy domain, protrude and make up the antigen-binding site (1, 2). Five of these CDR loops generally adopt only a limited number of backbone conformations, so-called canonical structures (reviewed in Ref. 3), which are determined by the lengths of the loops and by the presence of specific key residues. The antigen specificity of the binding site is mainly determined by the sequence and conformation of these CDR loops.

Antibody diversity is generated by the imprecise recombination of two or three sets of germline gene segments and by the combination of different heavy and light domains (4). The diversity is further increased by the process of somatic hypermutation (5) and by receptor editing and revision (6). As the germline variable gene repertoire encodes a rather limited number of CDR loop lengths (IMGT, the international ImMunoGeneTics data base, Ref. 7), the number of observed canonical structures is similarly limited. However, it was recently discovered that B cells evolve the genes encoding immunoglobulin V domains not only by nucleotide substitution but also through an additional mechanism of insertion and deletion of nucleotides during the hypermutation process (8–11). This mechanism has the potential to expand the available repertoire of loop lengths and conformations if the insertions and deletions involve entire codons and occur at positions in the sequence that can tolerate such modifications. A number of examples of seemingly functional insertions and deletions in the CDR of both the heavy and light domains of human antibodies have in fact been encountered lately (Refs. 8 and 12 and references therein). Furthermore, we have recently discovered that human IGHV[2] germline genes carry features in CDR1 and CDR2 that make these regions particularly prone to deletions of entire codons (12).

The occurrence of insertions and deletions in antibody V genes is not only of fundamental interest but is also of biotechnological importance. It has been known for some time that the topography of the antigen-binding site is related to the size of the antigen (13–15). Three different types of binding sites have been described: cavity, groove, and planar, which roughly correspond to hapten, peptide, and protein, respectively. This relationship has been further investigated by Vargas-Madrazo *et al.* (16), who have described a correlation between the length of the CDR loops and the antigen recognized. According to these findings, cleft-like binding sites that recognize small molecules are created by long loops (especially the CDRH2 and L1 loops), whereas planar-binding sites that are specific for large molecules are formed by short loops. In other words, by modifying the loop lengths of an antibody-binding site, it may thus be possible to design antibodies optimally suited for recognition of a particular class of antigen. Lamminmäki *et al.* (17) have in fact used this approach to modify a murine antibody specific for 17β-estradiol. They introduced additional residues into CDR2 of the heavy domain and were able to improve the recognition of the antigen. This improvement was suggested to be the result of a deeper binding site, created through the extension of CDRH2, which better accommodated the hapten (17).

Despite the establishment of insertions and deletions as naturally occurring modifications of antibody sequences and the use of amino acid insertions for antibody engineering, little is still known about the functional consequences of such mod-

* This study was supported by Biofavent Therapeutic AB, the Swedish Research Council, and the Crafoord Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 46-46-222-1322; Fax: 46-46-222-4200; E-mail: mats.ohlin@immun.lth.se.

[1] The abbreviations used are: V, variable; L, light; H, heavy; CDR, complementarity-determining region; scFv, single chain variable region fragment; FITC, fluorescein isothiocyanate; BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay; DSC, differential scanning calorimetry; PBS, phosphate-buffered saline.

[2] The immunoglobulin gene names used in this report are according to the official IMGT/HUGO nomenclature (IMGT, the international ImMunoGeneTics database, Ref. 7).

ifications. We have therefore created single-codon insertions and deletions as well as more complex modifications in the CDR of two human antibody single chain V region fragments (scFv) specific for a peptide and a hapten, respectively, and investigated the effects on antigen recognition, thermal stability, and protein folding. Our results demonstrate that single amino acid insertions in both CDRH1 and H2 and deletions in CDRH2 are usually well tolerated and permit production of folded proteins despite the fact that the modified loops carry amino acids that are disallowed at key residue positions in canonical loops of the corresponding length or do not take on a characteristic length of a known canonical structure. Modifications of this kind are in other words an efficient mode of expanding antibody sequence and structure space beyond what is encoded by the germline gene repertoire, which may enable targeting of novel or otherwise poorly immunogenic antigens.

EXPERIMENTAL PROCEDURES

Antibody Frameworks—The framework sequences encoding the anti-cytomegalovirus scFv AE11F and the anti-fluorescein isothiocyanate (FITC) scFv FITC8 have been described elsewhere (18–20). The cloning and production of the AE11F and AE11F/3-20L1 scFv in *Pichia pastoris* have also been described (21).

Creation of Insertion and Deletion Variants—Mini-libraries of scFv genes carrying codon insertions at various positions were created by the use of overlap extension PCR with degenerate primers that introduced NNK codons. Variants with a deletion were similarly created with primers lacking one codon. The AE11F based variants carrying CDRH1 sequences derived from the IGHV4 subgroup were created using the CDR-shuffling technique (22) essentially as described previously (21, 23).

Production and Purification of scFv Variants—The FITC8 scFv and all variant scFv were cloned into the pPICZα vector (Invitrogen) with C-terminal FLAG sequences (24) and produced in *P. pastoris* as described previously (21). The mini-libraries encoding AE11F and FITC8 variants were screened for scFv production or antigen binding according to the colony lift assay by McGrew *et al.* (25). Briefly, transformed *P. pastoris* colonies were lifted onto cellulose acetate filters (Pall Gelman Sciences, Ann Arbor, MI) and were grown on top of nitrocellulose filters, which were placed on methanol-containing plates. After 48 h of induction, scFv bound to the nitrocellulose filters were detected by a combination of anti-FLAG M2 antibody (Sigma) and rabbit anti-mouse Ig/horseradish peroxidase conjugate (DAKO A/S, Glostrup, Denmark) or FITC-biotin (Sigma) and streptavidin/horseradish peroxidase conjugate (DAKO A/S) using the ECL Plus™ Western blotting detection reagents (Amersham Biosciences) according to the manufacturer's recommendations. Single colonies were also picked and grown in liquid cultures to enable further characterization of the antigen binding properties (see below). In addition, a number of scFv variants were produced at a larger scale and purified as monomers. The AE11F-based variants were purified essentially as described previously (21), whereas the FITC8-based variants were purified by affinity chromatography on a Sepharose resin with FITC-conjugated bovine serum albumin (BSA) (kindly provided by Dr. B. Jansson, Biolvent Therapeutic AB, Lund, Sweden) followed by gel filtration as before.

Analysis of Antigen Recognition—The reactivity of the scFv variants with different antigens, both as crude expression supernatants and as purified monomers, was analyzed by enzyme-linked immunosorbent assay (ELISA) and by using the BIAcore technology (BIAcore AB, Uppsala, Sweden). The AE11F-based clones were tested on BSA, ovalbumin, streptavidin, and a biotinylated peptide that mimics the viral epitope (21) bound via streptavidin and the FITC8-based clones on BSA, streptavidin, FITC-BSA, FITC-biotin (bound via streptavidin), and a number of irrelevant BSA-coupled haptens obtained from Sigma or Biosearch Technologies Inc. (Novato, CA). The ELISA was performed according to standard protocols with anti-FLAG M2 (Sigma) and rabbit anti-mouse immunoglobulin/horseradish peroxidase conjugate (DAKO) to detect bound scFv. The BIAcore measurements and the calculation of the reaction rate kinetics were performed essentially as described previously (21).

Differential Scanning Calorimetry (DSC)—DSC measurements were performed using a VP-DSC from Microcal Inc. (Northampton, MA) in the temperature range 20–90 °C at a heating rate of 60°/h. All measurements were performed in phosphate-buffered saline (PBS), pH 7.4, containing 0.02% sodium azide at protein concentrations between 0.1 and 0.2 mg/ml with PBS in the reference cell. Prior to protein *versus* PBS measurements, PBS *versus* PBS scans were performed.

CD Spectroscopy—CD spectra were recorded on a J-720 spectropolarimeter (Jasco Inc., Easton, MD) in a 2-mm cuvette at a protein concentration of 0.1 mg/ml in 50 mM sodium phosphate, pH 7.4. Each sample was scanned two to eight times from 250 to 200 nm at a scan speed of 10 nm/min, a resolution of 1 nm, a bandwidth of 1 nm, and a sensitivity of 20 millidegrees, and the scans were combined to produce the final spectrum. Data are presented as mean residue molar ellipticity, which was calculated using the mean residue weight of each scFv.

Sequencing and Canonical Structure Classification—The nucleotide sequences of the variant scFv clones were determined by automated DNA sequencing as described elsewhere (26) after isolation of the templates by direct PCR on *P. pastoris* colonies using vector-specific primers. In the case of the CDRH1-grafted clones, the origin of the CDR was determined using the IMGT/V-QUEST alignment tool at IMGT, the international ImMunoGeneTics data base (imgt.cines.fr and Ref. 7). All sequences were defined and numbered in accordance with the IMGT nomenclature and unique numbering (7). Complete sequences of the variant scFv from this study can be found in GenBank™ under accession codes AF543317–AF543349. The canonical structure classification was performed using the software implemented on the Antibodies – Structure and Sequence server (www.bioinf.org.uk/abs/chothia.html and Ref. 27).

RESULTS

The scFv Frameworks—The parent antibody frameworks used in this study are both of human origin although there are differences in the way they were obtained. The AE11F scFv was derived from a monoclonal antibody isolated from a cytomegalovirus-seropositive blood donor (18, 19). It originates from the *IGHV3-30* and *IGKV3–11* genes, which both have acquired a number of mutations (21). This scFv recognizes both intact glycoprotein B from cytomegalovirus and peptides mimicking the AD-2 epitope (21, 28). The hapten (FITC)-specific scFv FITC8 was derived from a synthetic scFv library, which had been constructed by shuffling of human CDR sequences into a single framework consisting of the human *IGHV3–23* and *IGLV1–47* genes (20). The CDR sequences utilized by this scFv originate from *IGHV3–7* and *IGHV3–23* in the case of CDRH1 and CDRH2, *IGLV1–40* and *IGLV1–40* or *IGLV1–50* in the case of CDRL1 and CDRL2, and *IGLV1–47* in the case of CDRL3. Except for the CDRL1 loop, which is one residue longer than the *IGLV1–47* germline length, the CDR loops of the FITC8 scFv are of the same length as the loops normally encoded by the framework genes. As the structures of the two scFv have not been determined, the loop structures are unknown. However, by analyzing the deduced amino acid sequences using the tools at the Antibodies – Structure and Sequence server (27), the most similar of the observed canonical classes were identified (Table I).

Single-codon Insertions and Deletions—To determine the capability of the two antibody frameworks to tolerate length modifications in the CDR loops, we made single-codon insertions in CDRH1 and CDRH2 and a single-codon deletion in CDRH2. The modifications involved insertions after positions 31–33 in CDRH1, insertions after positions 57 and 58 in CDRH2, and a deletion at position 58 in CDRH2 (Fig. 1). All modifications were introduced at positions corresponding to the apices of the loops, *i.e.* the positions where the natural length variation occurs (31). A study of the IGHV germline gene repertoire has shown that these parts of the CDR carry repetitive sequence tracts, which naturally target them with deletions (and possibly also insertions) during the hypermutational process (12). Residues in these regions have also been shown to frequently make contact with the antigen in known antibody-antigen complexes (15), suggesting that modifications at the above mentioned positions will result in an expansion of structure space that is relevant for antigen recognition.

Libraries of scFv clones producing different insertion vari-

TABLE I

Examples of various single-codon modifications in scFv clones based on the AE11F and FITC8 frameworks, the canonical class belonging of the CDR loops, reactivity of the scFv with the original antigens as determined by ELISA or by BIAcore measurements, and the unfolding temperature of selected clones as determined by DSC

Modification refers to the nature of the changes in loop length; Ins indicates insertion, and Del indicates deletion. Numbering is according to the IMGT unique numbering (7). Canonical class indicates the combination of canonical structures of CDRH1, H2, and L1 as determined by automatic canonical structure classification (27). The altered canonical structure is indicated in *bold*. Antigen recognition: −, negative; ±, weakly positive; +, positive; + +, strongly positive.

scFv clone	Modification	Canonical class	Antigen recognition	Unfolding temperature
		H1-H2-L1		°C
AE11F	Original sequence	1-3-2	+ +	62.4
ASV18	Ins Pro-31A	2-3-2	+	
ASV19	Ins Asn-31A	2-3-2	+	
ASV43	Ins Arg-31A	2-3-2	+	
ASV15	Ins His-32A	2-3-2	+	
ASV37	Ins Ile-32A	2-3-2	±	
ASV39	Ins Phe-32A	2-3-2	+	
ASV02	Ins Phe-33A	2-3-2	±	
ASV35	Ins Asn-33A	2-3-2	−	
ASV07	Ins Lys-57A	1-U*-2	+ +	63.0
ASV08	Ins Ile-57A	1-U-2	+	
ASV28	Ins Thr-57A	1-U-2	+ +	
ASV05	Ins Glu-58A	1-U-2	+ +	
ASV10	Del Val-58	1-1-2	±	61.9
AE11F/3–20L1	Ins CDRL1[a]	1-3-**F**[b]	+	62.3
		H1-H2		
FITC8	Original sequence	1-3	+ +	63.4
FSV71	Ins Ser-31A	2-3	+	
FSV73	Ins His-31A	2-3	±	
FSV76	Ins Arg-31A	2-3	+	
FSV81	Ins Asn-32A	2-3	+	
FSV84	Ins Pro-32A	2-3	+ +	61.7
FSV85	Ins Arg-32A	2-3	+	
FSV91	Ins Leu-33A	2-3	+	
FSV93	Ins His-33A	2-3	+	
FSV96	Ins Tyr-33A	2-3	+	
FSV51	Ins Ser-57A	1-U	+ +	
FSV52	Ins Ala-57A	1-U	+ +	
FSV56	Ins Leu-57A	1-U	+ +	
FSV43	Ins Thr-58A	1-U	+ +	61.1
FSV46	Ins Arg-58A	1-U	±	
FSV61	Del Gly-58	1-1	+	63.9

[*] U indicates that the canonical structure of the created loop length is currently unknown.

[a] See text for details regarding the insertion.

[b] The automatic canonical class algorithms failed to unambiguously predict a structure for the CDRL1 loop of this scFv. Similarities in length and sequence with Fab 1f7 (PDB entry 1fig) suggest that the loop belongs to canonical structure class 6 (30).

ants were screened directly by the use of a colony lift assay (25). This analysis showed that ~95% of the clones based on the FITC8 framework had retained their specificity for FITC (data not shown). The libraries based on the AE11F framework were screened for the production of FLAG-carrying proteins, and a similar ratio of clones positive for scFv production was obtained (data not shown). Both positive and negative clones from each library were sequenced to determine the nature of the modifications, and the analysis showed that a wide range of amino acids was inserted at the intended positions. To determine the effect of these length modifications on the structure of the targeted loops, the most similar canonical structures were identified by the automatic canonical structure classification (27). A number of examples from each insertion library and the deletion variants are presented in Table I.

As the AE11F-based libraries were only tested for the production of FLAG-tagged proteins, they had to be characterized further to determine whether the scFv were functionally folded. This was done by analyzing the antigen-binding properties of the modified clones. Although changes in loop structure may be associated with a loss of antigen recognition, specific recognition of an antigen will confirm that the polypeptide chain is correctly folded as this is a requirement for it to function as a framework for the antigen-binding site. Analysis of expression supernatants of randomly picked clones (including the deletion variants) by ELISA or by using the BIAcore technology confirmed the above finding that the majority of the FITC8-based clones recognized the original antigen. Importantly, this analysis showed that most of the AE11F-based clones had also retained their specificity for the original viral antigen (Table I). Furthermore, when tested for binding to a number of irrelevant antigens (see "Experimental Procedures"), none of the clones displayed any cross-reactivity (data not shown), demonstrating that the modified scFv clones retained a high degree of specificity for the original antigens and therefore likely also assumed a correct immunoglobulin fold.

A number of clones of each specificity, chosen to exemplify the different modifications, were produced at a large scale to study the interaction with the original antigens in detail and determine the stability of the purified proteins. BIAcore measurements with the purified monomers of the ASV07, ASV10, ASV35, FSV43, FSV61, and FSV84 clones confirmed the previously obtained results with crude expression supernatants (Table I and Fig. 2). Furthermore, evaluation of the reaction rate kinetics with the original antigen showed that the modifications did not affect the dissociation rates of the FITC8-based clones to any greater extent (Fig. 2*B*). The thermal stability of the purified monomers was determined by DSC, and all tested clones displayed unfolding temperatures very similar to the parent scFv (Table I), further verifying that the IGHV3-derived antibody frameworks tolerate single-codon insertions and deletions in CDRH1 and H2 very well.

The Journal of Biological Chemistry

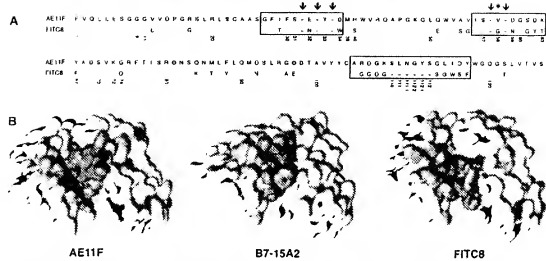

FIG. 1. **Sequences and structures of the scFv frameworks used for production of insertion and deletion variants.** *A*, alignment of the deduced amino acid sequences of the heavy V domains of the AE11F and FITC8 scFv. CDR-IMGT are *boxed*, and the location of the insertions and the deletion made in this study are indicated by *arrows* and an *asterisk*, respectively. Amino acid *numbering* according to the IMGT unique numbering is shown *below* the sequences. *B*, location of the affected sequences as indicated on a structure model of AE11F, which was generated using the WAM algorithm (29), a determined structure of the protein-specific antibody B7-16A2 (Protein Data Bank entry 1ngk), which originates from a highly related IGHV gene and has a CDRH3 of the same length as AE11F, and a structure model of FITC8 (20). CDRH3 is shown in *red*, whereas residues immediately *adjacent* to the single-codon insertions and the deletion made in this study are *highlighted* in *blue* (residues 31–34 in CDRH1) and *green* (residues 57–59 in CDRH2), respectively.

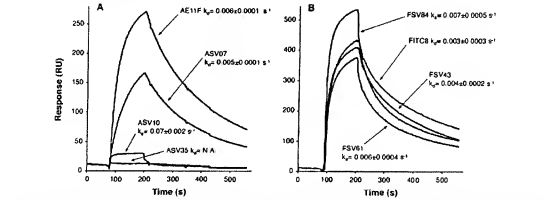

FIG. 2. **The single-codon modifications did not affect the reaction rate kinetics of the FITC8-based scFv variants to any greater extent.** Representative BIAcore sensorgrams of AE11F-based scFv and FITC8-based scFv analyzed on streptavidine-bound viral peptide (*A*) or streptavidine-bound FITC (*B*), respectively. Dissociation rate constants (k_d) were calculated from multiple measurements and are presented as the mean value ± S.E. *N.A.*, not applicable; *RU*, resonance units.

As insertions and deletions have been demonstrated to occur naturally in both heavy and light domain V genes (8), we decided to extend this study and also evaluate the stability of a previously produced AE11F-based scFv variant with an insertion in CDRL1 (AE11F/S-20L1) (21). The modified CDRL1 of this scFv is identical, except for an additional serine residue, to the germline gene from which AE11F originates. This clone has also been demonstrated to recognize both the epitope-mimicking peptide and intact, recombinant glycoprotein B, albeit with a lower affinity than the affinity matured AE11F scFv (21, 32). The thermal stability of the AE11F/S-20L1 scFv was determined as before after purification of monomeric scFv, and the unfolding temperature was found to be similar to that of the original scFv (Table I), thus indicating that not only heavy but also light domain CDR tolerate modifications of this nature well.

Grafting of CDRH1 Loops from Distantly Related IGHV Genes—As all of the insertions and deletions described so far were introduced at the tips of the hypervariable loops, the parts of the immunoglobulin fold that best can be expected to accommodate such modifications, we decided to introduce more extensive modifications to investigate the effect of such changes of antibody sequence and structure. These modifications were introduced into and immediately adjacent to CDRH1 of the AE11F framework by the CDR-shuffling technique (22) using CDR sequences isolated from activated human B cells. Sequences originating from the IGHV4 subgroup were chosen for

TABLE II

Deduced amino acid sequences, germline gene origin, and canonical structure class belonging of the CDRH1 loops of the AE11F scFv and the CDRH grafted variants of this

Amino acid sequences are aligned and numbered in accordance with the IMGT unique numbering (7) and gaps thereby introduced are indicated by *dashes*. Amino acids that are part of the CDR1-IMGT (7) are *underlined*. *Dots* indicate identity with the AE11F sequence. Canonical structures were determined by automatic canonical structure classification (27).

scFv clone	Amino acid sequence	Germline origin	Canonical structure
	2 3 33334 4 2 0 45690 4		
AE11F	SCAASGFIFSRYD--MSWVRQ	*IGHV3-30*	1
E3	...V..GSI.SGGYWS.I..	*IGHV4-31*	3
E6	...V..YSI.SGYY-WG.I..	*IGHV4-b*	2
E10	...V..GSI.S.Y--WS.I..	*IGHV4-59*	1
E11	...V..GSI.G.H--WS.I..	*IGHV4-34*	1
E14	...V..GSI.SGGYSWS.I..	*IGHV4-30-2*	3

the grafting as these are only distantly related to the IGHV3 CDR and therefore allow for a higher degree of variability. In addition, genes from the IGHV4 subgroup encode loops of different lengths than genes from the IGHV3 subgroup, including loops of the same length as the ones created by the single-codon insertions in CDRH1, thus enabling a comparison with these modifications. Sequencing of randomly picked clones showed that seemingly functional, *i.e.* in-frame and without stop codons, IGHV3 genes carrying IGHV4-derived CDRH1 sequences were obtained (Table II). However, when analyzing crude expression supernatants of the constructs, it was found that all of the clones had lost the original antigen specificity and instead acquired a polyreactive character (Fig. 3).

To further investigate this polyreactive nature of the CDRH1-grafted clones, two of them, E3 and E6, were produced at a larger scale and purified as monomers to enable structural characterization. These two clones were chosen based on the presence of loop lengths different from the one used by the parent antibody (Table II). As judged by analytical gel filtration, these clones also gave rise to proteins that behaved as scFv monomers (data not shown). The overall secondary structure was determined by CD spectroscopy and was compared with the results obtained with other monomeric scFv. As shown in Fig. 4, the spectra of both of the CDRH1-grafted clones displayed a strong negative signal near 200 nm, which is indicative of unordered polypeptides (33). For a comparison, the spectra of both the parent scFv and the FITC8 scFv displayed a weak negative signal near 217 nm, which is characteristic of the β-sheet conformation of antibody domains (Fig. 4). The same result was also obtained with clones carrying single-codon modifications, such as the AE11F/3-20L1 and the FSV43, which gave rise to nearly identical spectra as the parent scFv (data not shown). When analyzed by DSC, no unfolding temperatures could be determined for either of the E3 or E6 scFv, suggesting that the proteins already were in an, at least partly, unfolded state. Thus, by inserting these only distantly related CDR sequences into the IGHV3 framework, the boundaries that define a stable immunoglobulin fold had apparently been exceeded.

DISCUSSION

Insertions and deletions of nucleotides have recently been shown to be an additional mechanism whereby immunoglobulin V region genes are evolved (8–11) and which may expand the available repertoire of antibody hypervariable loop lengths and structures. Although sequence modifications of this kind, especially insertions, have also been exploited in antibody engineering, knowledge about the effects of these modifications on protein stability and antigen recognition is still limited. Such factors are critical as they determine the success of this mode of molecular evolution, whether employed by nature or by the molecular engineer. To study the functional consequences of both insertions and deletions in the CDR of human antibodies, we have here made single-codon insertions and deletions as well as more extensive modifications in the CDR of two antibody fragments with different specificities and assessed the thermal stability and the antigen binding properties of the resulting proteins.

The single-codon modifications were well tolerated by the two scFv frameworks as determined by the thermal stability measurements and the high ratio of functional clones despite the fact that they created both loop lengths that do not occur normally within the human IGHV3 subgroup and combinations of loop lengths that do not exist in the human germline repertoire. Insertion of one residue in CDRH2 of the two scFv studied here creates a loop length (CDR2-IMGT length 9 amino acids) that is not naturally encoded by any IGHV genes except for the only member of the IGHV6 subgroup (7). This loop length has been predicted to have its own distinct conformation (canonical structure 5, Ref. 31), but as no immunoglobulin encoded by this gene has been structurally determined, this canonical structure has not been defined. The insertion of one residue in CDRH1 produces a loop length (CDR1-IMGT length 9 amino acids) that occurs naturally within the human IGHV4, but not the IGHV3 subgroup, and which could correspond to canonical structure 2 as judged by the automatic canonical structure classification. This coexistence of canonical structure 2 in CDRH1 with canonical structure 3 in CDRH2 (Table I) does not occur naturally within the human IGHV germline repertoire, although it has been observed in hypermutated antibodies with insertions in CDRH1 (8). In addition, the structure classification also revealed that a large number of the key residue requirements for canonical structure 2 were not fulfilled (27), *i.e.* the thus modified CDRH1 loops either take on structures not covered by the described canonical structures or adopt the observed structure corresponding to this loop length despite the presence of a large number of disallowed amino acids at key residue positions. Irrespective of the circumstances, the insertions in CDRH1 seem to, like the rest of the single-codon modifications, give rise to scFv that are correctly folded and stable.

The fact that the loop lengths that were created by the single-codon insertions are not part of the IGHV3-encoded repertoire does not mean that they are completely unnatural in the context of an IGHV3 framework. Apparently functional antibodies belonging to the IGHV3 subgroup with insertions in CDRH1 and CDRH2 leading to CDR-IMGT loop lengths of 9 amino acids have in fact been described by others (8, 34, 35). As the deletions at position 58 in CDRH2 of both scFv give rise to loop lengths that are used by other members of the IGHV3 subgroup, it is not entirely unexpected that these modifications

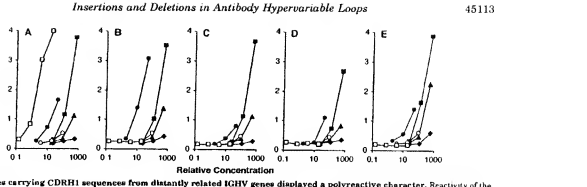

FIG. 3. Clones carrying CDRH1 sequences from distantly related IGHV genes displayed a polyreactive character. Reactivity of the AE11F (□), E3 (■), E6 (○), E10 (▲), E11 (◆), and E14 (●) scFv with streptavidin-bound viral peptide (*A*), streptavidin (*B*), BSA (*C*), ovalbumin (*D*), and uncoated polystyrene wells (*E*), as determined by ELISA. Relative concentrations of the expression supernatants were estimated by immunoblotting. The coefficient of variation was below 10% for the whole data set.

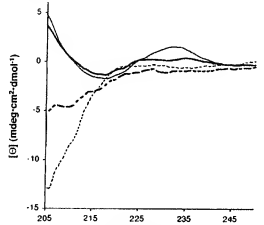

FIG. 4. CD spectroscopy indicated an unordered folding of the CDRH1-grafted clones. CD spectra of purified monomers of the AE11F (*thick solid line*), P7C8 (*thin solid line*), E3 (*thick broken line*), and E6 (*thin broken line*) scFv in 50 mM sodium phosphate, pH 7.4.

are tolerated by the scFv frameworks studied here. Furthermore, in a previous study, we have found that single-codon deletions, some of which have also been shown to be functional, occur in antibodies belonging to the IGHV3 subgroup at or immediately adjacent to position 58 (12). The single-codon modifications of antibody sequence space we have presented here are in other words highly representative of changes that may occur naturally as a consequence of the somatic hypermutation process.

As some of the single-codon insertions produced loop lengths found in antibodies belonging to the IGHV4 subgroup, we decided to investigate the possibility of using CDRH1 sequences originating from this subgroup to diversify the AE11F scFv. This approach resembles evolution through receptor revision, which occurs *in vivo* (36, 37) and has also been shown to provide a selection advantage *in vitro* (38). However, grafting of CDRH1 loops of different lengths from the IGHV4 subgroup into the IGHV3 framework used by the AE11F scFv resulted not only in a loss of the original antigen specificity but also in the acquisition of a polyreactive character, even when not having been put through a potentially denaturing purification process (39), by the thus modified scFv clones (Fig. 3). This polyreactivity is most likely due to a destabilized or inappropriately folded V domain, as demonstrated by the CD spectra of two of the clones (Fig. 4). Destabilizing effects of loop grafting into an antibody framework have been reported previously (40), but in that particular case, the grafted sequences were totally unrelated to antibody hypervariable loops. The use of naturally occurring CDR sequences for grafting into immunoglobulin frameworks often ensures that the inserted loops are optimally functional as they have been proofread and selected for functionality during the formation of the B cell receptors. Our data show, however, that the functionality of the grafted loops also depends on the framework they are inserted into even if they are natural immunoglobulin sequences. The reason for the observed effects probably lies in the differences in certain key residues between the IGHV3 and IGHV4 frameworks. In fact, many of the amino acids that differ between the original AE11F sequence and the grafted sequences are residues that are used to define the canonical structures (27, 31). In addition, Tramontano *et al.* (41) have shown that framework residue 89 of the heavy V domain packs against residues in both CDRH1 (position 30) and CDRH2 (position 58) and that it is an important determinant of the conformation of the CDRH2 loop. A subsequent mutational study has also shown that the nature of this residue determines the binding characteristics of an antibody by influencing the conformation of the heavy chain CDR loops (42). The AE11F framework has, like all unmutated antibodies belonging to the IGHV3 subgroup, an Arg at position 89, whereas all genes belonging to the IGHV4 subgroup, from which the CDRH1 sequences were obtained, encode a Val residue at this position in their germline configurations. The larger, charged Arg possibly causes clashes with the IGHV4-derived residues in and adjacent to CDRH1, which leads to an improper fold and poor stability of the resulting scFv product.

In conclusion, we demonstrate here that single amino acid insertions in both CDRH1 and H2 and deletions in CDRH2, which are highly representative of modifications that occur naturally in regions of the hypervariable loops known to be involved in antigen contact (15) during the maturation of B cell receptors, are well tolerated and permit production of stably folded proteins. This is true despite the fact that the thus modified loops do not fulfill the key residue requirements for canonical loops of the corresponding length or are of a length not associated with a known canonical structure (27). This demonstrates the plasticity of antibody V domain frameworks belonging to the important IGHV3 subgroup, which makes up a large fraction of all human antibodies (43), and its capacity to tolerate modifications that expand sequence and structure space beyond the limits set by the germline-encoded diversity.

Based on the similarities with naturally occurring alterations of loop lengths, our results with insertions and deletions in CDRH1, H2, and L1 of the antibody fragments used in this study, and work on an unrelated scFv with a three-amino acid insertion at the beginning of CDRH1 (10),[3] our conclusion is that both insertions and deletions can be efficiently utilized in antibody engineering to expand the structural space available to human antibodies as long as attention is paid to key residues in the framework (41). As demonstrated by previous studies on murine antibodies, this approach can be used for improving already existing specificities (17, 44). However, analogously with the correlation between CDR loop lengths and the antigen recognized (16), it is conceivable that it may also be utilized for the construction of antibody libraries specific for a particular class of antigens such as haptens, peptides, or large molecules. Finally, we hypothesize that introduction of novel loop lengths and combinations of loop lengths not encoded by the germline repertoire may also enable the targeting of poorly immunogenic or previously unrecognized antigens and epitopes as entirely new regions of antibody structure space are explored by this mode of sequence diversification.

Acknowledgments—We thank Ola Jakobsson and Micael Oswald for technical assistance.

REFERENCES

1. Amzel, L. M., and Poljak, R. J (1979) *Annu. Rev. Biochem.* **48**, 961–997
2. Padlan, E. A. (1994) *Mol. Immunol.* **31**, 169–217
3. Al-Lazikani, B., Lesk, A. M., and Chothia, C. (1997) *J. Mol. Biol.* **273**, 927–948
4. Tonegawa, S (1983) *Nature* **302**, 575–581
5. Berek, C., and Milstein, C (1987) *Immunol. Rev.* **96**, 23–41
6. Nemazee, D., and Weigert, M. (2000) *J. Exp. Med.* **191**, 1813–1817
7. Lefranc, M. P. (2001) *Nucleic Acids Res.* **29**, 207–209
8. de Wildt, R. M., van Venrooij, W. J., Winter, G., Hoet, R. M., and Tomlinson, I. M. (1999) *J. Mol. Biol.* **294**, 701–710
9. Cosgarea, T., Nkon, U., and Küppers, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2463–2468
10. Ohlin, M., and Borrebaeck, C. A. K. (1998) *Mol. Immunol.* **35**, 233–238
11. Wilson, P. C., de Boutaillier, O., Liu, Y. J., Potter, K., Banchereau, J., Capra, J. D., and Pascual, V. (1998) *J. Exp. Med.* **187**, 59–70
12. Lantto, J., and Ohlin, M. (2002) *J. Mol. Evol.* **54**, 345–353
13. Wilson, I. A., and Stanfield, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 113–118
14. Webster, D. M., Henry, A. H., and Rees, A. R. (1994) *Curr. Opin. Struct. Biol.* **4**, 123–129
15. MacCallum, R. M., Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **262**, 732–745
16. Vargas-Madrazo, E., Lara-Ochoa, F., and Almagro, J. C. (1995) *J. Mol. Biol.* **254**, 497–504
17. Lamminmäki, U., Pauperio, S., Westerlund-Karlsson, A., Kaartinen, J., Virtanen, P. L., Lövgren, T., and Saviranta, P. (1999) *J. Mol. Biol.* **291**, 589–602
18. Ohlin, M., Sundqvist, V. A., Mach, M., Wahren, B., and Borrebaeck, C. A. K. (1993) *J. Virol.* **67**, 703–710
19. Ohlin, M., Owman, H., Mach, M., and Borrebaeck, C. A. K. (1996) *Mol. Immunol.* **33**, 47–56
20. Söderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Åberg, A. M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., Danielsson, L., Carlsson, R., and Borrebaeck, C. A. K. (2000) *Nat. Biotechnol.* **18**, 852–856
21. Lantto, J., Lindroth, Y., and Ohlin, M. (2002) *Eur. J. Immunol.* **32**, 1659–1669
22. Jirholt, P., Ohlin, M., Borrebaeck, C. A. K., and Söderlind, E. (1998) *Gene (Amst.)* **215**, 471–476
23. Lantto, J., Jirholt, P., Barmos, V., and Ohlin, M. (2002) *Methods Mol. Biol.* **178**, 303–316
24. Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Cerretti, D. P., Urdal, D. L., and Conlon, P. J (1988) *Biol Technology* **6**, 1204–1210
25. McGrew, J. T., Lemke, D., Dell, B., Klinke, R., Krentz, D., Wee, S. F., Abbott, N., Armitage, R., and Harrington, K. (1997) *Gene (Amst.)* **197**, 193–200
26. Jirholt, P., Strandberg, L., Jansson, B., Kramborvita, E., Söderlind, E., Borrebaeck, C. A. K., Carlsson, R., Danielsson, L., and Ohlin, M. (2001) *Protein Eng.* **14**, 67–74
27. Martin, A. C., and Thornton, J. M. (1996) *J. Mol. Biol.* **263**, 800–815
28. Lantto, J., Fletcher, J. M., and Ohlin, M. (2002) *J. Gen. Virol.* **83**, 2001–2005
29. Whitelegg, N. R., and Rees, A. R. (2000) *Protein Eng.* **13**, 819–824
30. Tomlinson, I. M., Cox, J. P., Gherardi, E., Lesk, A. M., and Chothia, C. (1995) *EMBO J.* **14**, 4628–4638
31. Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B., and Winter, G. (1992) *J. Mol. Biol.* **227**, 799–817
32. Lantto, J., Fletcher, J. M., and Ohlin, M. (2002) *Virology*, in press
33. Woody, R. W. (1995) *Methods Enzymol.* **246**, 34–71
34. Brezinschek, H. P., Foster, S. J., Brezinschek, R. I., Dorner, T., Domiati-Saad, R., and Lipsky, P. E. (1997) *J. Clin. Invest.* **99**, 2488–2501
35. Hoppe, S. M., Herman, C., Bakkus, M. H., Simonick, J., Schots, R., and Thielemans, K. (1999) *Br. J. Haematol.* **107**, 625–640
36. Wilson, P. C., Wilson, K., Liu, Y. J., Banchereau, J., Pascual, V., and Capra, J. D. (2000) *J. Exp. Med.* **191**, 1881–1894
37. Heb, K., Meffre, E., Albanosmo, E., Farber, A., Dines, D. Stein, P., Aanie, S. E., Furie, R. A., Jain, R. I., and Chiorazzi, N. (2000) *J. Exp. Med.* **192**, 1151–1164
38. Ellmark, P., Estéban, O., Furebrand, C., Malmborg Hager, A.-C., and Ohlin, M. (2002) *Mol. Immunol.* **39**, 249–256
39. McMahon, M. J., and O'Kennedy, R. (2000) *J. Immunol. Methods* **241**, 1–10
40. Helms, L. R., and Wetzel, R. (1995) *Protein Sci.* **4**, 2073–2081
41. Tramontano, A., Chothia, C., and Lesk, A. M. (1990) *J. Mol. Biol.* **215**, 175–182
42. Xiang, J., Sha, Y., Jia, Z., Prasad, L., and Delbaere, L. T. (1995) *J. Mol. Biol.* **253**, 385–390
43. de Wildt, R. M., Hoet, R. M., van Venrooij, W. J., Tomlinson, I. M., and Winter, G. (1999) *J. Mol. Biol.* **285**, 895–901
44. Pedersen-Sørsen, B., Viswanathan, M., and Margolies, M. N. (2002) *J. Immunol. Methods* **259**, 43–53

[3] J. Lantto and M. Ohlin, unpublished work